

上海交通大学

Cuckoo 性能测试 实验报告

学 院：电子信息与电气工程学院

学生姓名：杨景凯

学 号：520021910550

2022 年 5 月 14 日

目录	2
----	---

目录

1 背景介绍 [1]	3
1.1 哈希表	3
1.2 Cuckoo 哈希	3
2 测试	3
2.1 测试内容	3
2.2 测试方式	3
2.3 测试结果	4
3 结论	4

1 背景介绍 [1]

1.1 哈希表

哈希表是一种可以快速寻找目标键值对的方式，但是在多个键对应的哈希值相同时就会发生“冲突”。为了应对这种冲突，不同的哈希表会有不同的冲突应对策略——线性哈希在对应哈希值的基础上继续向下查找空位，并将其放在第一个空位上；链式哈希表则会将对应哈希值的所有键值对串成链表，查找时遍历链表。

1.2 Cuckoo 哈希

Cuckoo 采用了不同的设计来减少冲突情况下查找的代价。即每个键有且只有两个不同的哈希值，如果两个哈希值对应的位置存在空位就存放在其中。如果两个哈希值对应的位置不存在空位，那就需要转移对应位置中已有的键值对，直到所有键值对都落在它们两个哈希值之一对应的指定位置处。当然这种方法在键值对转移过程中可能会出现死循环，解决方法是路径检测加上重新哈希。检测循环路径的方法也比较简单，可以预先设定一个阈值，当循环次数或者递归调用次数超过设定阈值时，就可以认为产生了循环路径。

2 测试

2.1 测试内容

以哈希表容量为 100000 为例，在负载因子为 0.1, 0.5, 0.75, 1.0 的情况下，插入数据并测试 Cuckoo 哈希的查找时间（测试查找成功的情况即可）

2.2 测试方式

测试方式如下：

1. 先将总数据量为 $100000 \cdot \alpha$ (α 表示负载因子) 的数据顺序插入 Cuckoo 哈希表中，对于插入数据，我选取了简单的 $0 \sim 100000 \cdot \alpha$ 的数据。
2. 将所有数据（即 $0 \sim 100000 \cdot \alpha$ 的数据，共计 $100000 \cdot \alpha$ 个）进行搜索，每一数据搜索 1000 次，并除以 $1000 \cdot \alpha$ ，计算得到搜索任意一个数据的平均查找时间乘以 100000 的时间 (ms)。

2.3 测试结果

测试结果如下表与图所示：

负载因子	0.1	0.5	0.75	1
乘以 100000 后的平均查找时间 (ms)	0.45	0.42	0.425333333	0.424

表 1: Cuckoo 哈希表容量为 100000 时负载因子与平均查找时间 (ms) 关系表格

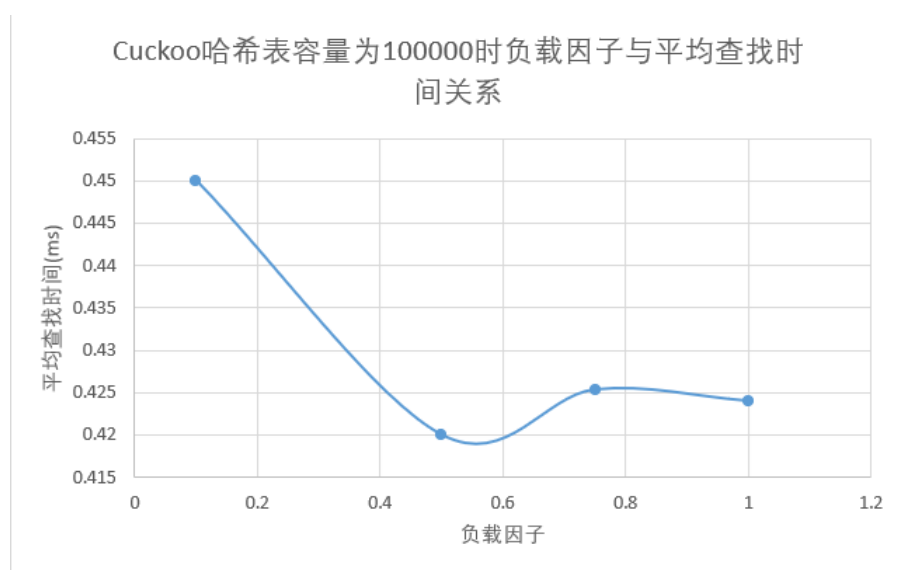


图 1: Cuckoo 哈希表容量为 100000 时负载因子与平均查找时间 (ms) 关系图

3 结论

观察表格与图片可以发现，Cuckoo 哈希表容量为 100000 时，负载因子与平均查找时间没有关系，不同负载因子下，对于 Cuckoo 哈希表的平均查找时间相近，故可以证明 Cuckoo 哈希表是一种优秀的查找数据的结构。

参考文献

- [1] Cuckoo 性能测试.pdf